Higher order quasi-Monte Carlo methods: A comparison

Dirk Nuyens and Ronald Cools

Department of Computer Science, K.U.Leuven, Belgium

Abstract. Quasi-Monte Carlo is usually employed to speed up the convergence of Monte Carlo in approximating multi-variate integrals. While convergence of the Monte Carlo method is $O(N^{-1/2})$, that of plain quasi-Monte Carlo can achieve near $O(N^{-1})$. Several methods exist to increase its convergence to near $O(N^{-\alpha})$, $\alpha > 1$, if the integrand has enough smoothness. We discuss two methods: lattice rules with periodization and higher order digital nets, and present a numerical comparison.

Keywords: Quadrature, Cubature, Quasi-Monte Carlo, Higher order of convergence, Lattice rules and sequences, Digital nets and sequences

INTRODUCTION

We are concerned with approximating a multi-variate integral by a simple cubature formula Q,

$$I(f) := \int_{[0,1]^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad \approx \qquad Q_N(f;\{\boldsymbol{x}^{(k)}\}_k,\{w^{(k)}\}_k) := \sum_{k=0}^{N-1} w^{(k)} f(\boldsymbol{x}^{(k)}),$$

where the integrand f has sufficient smoothness. A simple *Monte Carlo* estimator is given by

$$\frac{1}{N}\sum_{k=0}^{N-1} f(\boldsymbol{x}^{(k)}), \tag{1}$$

where the sample nodes $\boldsymbol{x}^{(k)}$ are i.i.d. uniform variates over $[0,1]^s$; it achieves a convergence of $O(N^{-1/2})$. The *quasi-Monte Carlo* method has exactly the same form as (1) but uses a deterministic set of sample nodes, called *low-discrepancy points*. Under certain conditions, its convergence is much quicker and approaches $O(N^{-1})$ (where here we have hidden the dependence on s). In this paper we are interested in low-dimensional integrals. A general reference on quasi-Monte Carlo is [1].

From classical quadrature we know that there exist quadrature formulae that achieve a convergence of $O(N^{-\alpha})$ for functions having α continuous derivatives. For multi-variate functions there are similar results: if the integrand function f is in a class H_s^α of functions having mixed dominating smoothness α, i.e., all

$$\frac{\partial^{\nu_1,\ldots,\nu_s} f(\boldsymbol{x})}{\partial x_1^{\nu_1}\cdots\partial x_s^{\nu_s}} \qquad \text{are of bounded variation (BV) for all } 0 \le \nu_i \le \alpha - 1, \tag{2}$$

then there exist quasi-Monte Carlo rules which achieve near $O(N^{-\alpha})$; this *bounded variation* is the s-variate Hardy and Krause variation, see, e.g., [1]. A sufficient, but unnecessary, condition for (2) is that $\partial^{\alpha s} f(\boldsymbol{x})/\partial x_1^\alpha \cdots \partial x_s^\alpha$ exists and is continuous on $[0,1]^s$.

We now present two methods to obtain a higher order of convergence: a classical approach using lattice rules with periodization, and a recent and promising technique of higher order digital nets.

METHOD 1: LATTICE RULES AND PERIODIZATION

The method of "good lattice rules" was developed for *periodic* functions, see, e.g., [2, 3, 5]. That is, in addition to (2), we also require the multi-variate analog of $f^{(\nu)}(1) = f^{(\nu)}(0)$, $0 \le \nu \le \alpha - 2$ to hold:

$$\left.\frac{\partial^{\nu_1,\ldots,\nu_s} f(\boldsymbol{x})}{\partial x_1^{\nu_1}\cdots\partial x_s^{\nu_s}}\right|_{x_j=1} = \left.\frac{\partial^{\nu_1,\ldots,\nu_s} f(\boldsymbol{x})}{\partial x_1^{\nu_1}\cdots\partial x_s^{\nu_s}}\right|_{x_j=0} \qquad \text{for all } j \text{ and } 0 \le \nu_i \le \alpha - 2. \tag{3}$$

Then there exist *generating vectors* $\boldsymbol{z} \in \mathbb{Z}^s$ such that the point set with points

$$\boldsymbol{x}^{(k)} := \frac{\boldsymbol{z}k \bmod N}{N}, \qquad \text{for } k = 0, \ldots, N-1,$$

achieves a convergence of $O(N^{-\alpha}(\log N)^{(s-1)\alpha})$. Such a rule is called a *lattice rule*. Such lattice rules can for example be constructed with the methods in [6, 7]. (We note that by considering special weighted classes of functions, the dependence on s can be removed, see, e.g., [8].)

Obviously, the requirement (3) is rather strong and most functions in real-world problems do not fulfill this. A particularly simple strategy exists to change a smooth function which fulfills (2) into a smooth periodic function which also adheres to (3) by a suitable variable substitution. This simple strategy is called *periodization*. We replace f by g, by a change of variables,

$$g(\boldsymbol{x}) = \varphi'(x_1) \cdots \varphi'(x_s)\, f(\varphi(x_1), \ldots, \varphi(x_s)),$$

where φ is a smooth increasing function for which $\varphi(0) = 0$, $\varphi(1) = 1$ and $\varphi^{(\nu)}(0) = \varphi^{(\nu)}(1)$ for $\nu = 1, \ldots, \alpha - 1$, such that $\int_{[0,1]^s} g(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{[0,1]^s} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$. We further assume φ' to be symmetric. Many such transformations have been developed, e.g., [3, 9, 10]. The equal weight rule (1) can now be reinterpreted to have weights:

$$\sum_{k=0}^{N-1} \underbrace{N^{-1}\, \varphi'(x_1^{(k)}) \cdots \varphi'(x_s^{(k)})}_{=:w^{(k)}}\, f(\varphi(x_1^{(k)}), \ldots, \varphi(x_s^{(k)})) = \sum_{k=0}^{N-1} w^{(k)}\, f(\varphi(\boldsymbol{x}^{(k)})). \tag{4}$$

The convergence of such a lattice rule with periodization is now the desired $O(N^{-\alpha}(\log N)^{(s-1)\alpha})$. However, as noted in [11], the error bound consists of two parts:

$$|I(g) - Q(g)| \le P_\alpha(\boldsymbol{z})\, V_\alpha(g), \tag{5}$$

where the first part is a measure of the quality of the lattice rule, attaining the claimed $O(N^{-\alpha}(\log N)^{(s-1)\alpha})$, and the second part, the "*variation*" of g, is dependent on the function. It is this second part which might be much worse compared to a similar bound for the non-periodic space which would just involve f. Further negative publicity was given to this method by studying its tractability in [12].

We end this exposition on the periodizing strategy by listing some positive and negative points. The main negative point is on the limited dimensionality that is possible. This was already noted in [3, 5] and more recently, in a different analysis, in [12], but for different reasons than we discuss here.

- – There are exponentially large differences in the weights $w^{(k)}$: By the conditions on φ it is necessary that φ' is small at the sides of $[0,1]$ while it has to be larger than 1 in the middle. Since the $w^{(k)}$ are of product form the weights vanish at the sides and grow in the middle of $[0,1]^s$, both exponentially in s. Even for a low dimensionality one quickly needs to resort to multi-precision calculations. This makes it practically very hard to evaluate (4).
- – Furthermore, in [12] it is shown that in some cases the sum of weights $\sum_{k=0}^{N-1} w^{(k)}$ can go to infinity, exponentially fast with s. The numerical stability of the rule (4) depends on this sum as the condition number w.r.t. perturbations in $f(\varphi(\boldsymbol{x}))$ of (4) can be bounded by $\sum_{k=0}^{N-1} |w^{(k)}|$, and here all $w^{(k)} > 0$. In contrast, the condition number of the mathematical problem of integrating f is just 1. In such a case the rule (4) may become numerically unstable.
- – Some of the periodization substitutions can be costly, but this is of course relative to the cost of evaluating the original function f. (As in [13] this can however be pre-calculated.)
- – Periodization is likely to increase the variation $V_\alpha(g)$ of the function exponentially in both α and s, which is a direct factor in the error bound (5) above.
- + The technique of periodization is particularly simple to implement. One could for example just change the implementation of f to incorporate a specific periodization.
- + There is a large choice of periodizing transformations, see, e.g., [3, 9, 10].
- + Lattice rules with high order of convergence can be easily constructed, see, e.g., [6]. An explicit generating vector for $\alpha = 3$ is given in [14].
- + The effectiveness has been demonstrated before, e.g., [13] and [15]. Particularly [13], a 2-dimensional automatic cubature routine, shows how to pre-compute the weights and transformed lattice points in higher precision to circumvent some of the drawbacks.

METHOD 2: HIGHER ORDER DIGITAL NETS

Given a "vector" of matrices $\boldsymbol{C} \in (\mathbb{F}_b^{n\times m})^s$, then the coordinates of the kth point of a digital net in base b, with $N = b^m$ points, are obtained by matrix vector products over $\mathbb{F}_b$, i.e.,

$$\boldsymbol{x}^{(k)} := \boldsymbol{C}\cdot k, \quad \text{for } k = 0,\dots,N-1, \quad \text{where } x_j^{(k)} = C_j \cdot k \simeq \begin{pmatrix} x_{j,1}^{(k)} \\ \vdots \\ x_{j,n}^{(k)} \end{pmatrix} = C_j \begin{pmatrix} k_0 \\ \vdots \\ k_{m-1} \end{pmatrix} \text{ over } \mathbb{F}_b, \text{ for } j = 1,\dots,s, \quad (6)$$

where the $x_{j,i}^{(k)}$ are interpreted as the digits of an n-digit base b rational in $[0,1)$, i.e., $x_j^{(k)} = \sum_{i=1}^{n} x_{j,i}^{(k)}\, b^{-i}$, and similarly $k = \sum_{i=0}^{m} k_i\, b^i$. If these *generating matrices*, $C_j \in \mathbb{F}_b^{n\times m}$, are square $m\times m$ matrices then one has the classical digital nets as described in [1]. The new techniques in [16] consider non-square matrices, where it is shown that for $n = \alpha m$, there exist generating matrices such that functions having mixed dominating smoothness α can be numerically integrated by (1) with a convergence of near $O(N^{-\alpha})$.

An explicit construction method based on existing (classical) digital nets was given in [16] as well. If one considers functions with mixed dominating smoothness α in s dimensions, then, as the basis of the construction, one can use an existing digital net of dimension αs and reassemble the generating matrices to obtain a higher order net. For such a higher order net in base b, which has b^m points, the worst-case error behaves like

$$O\Big(b^{-\alpha \max(0, m-(t'_{\alpha s}+\lfloor s(\alpha-1)/2\rfloor))}\Big), \quad (7)$$

where $0 \le t'_{\alpha s} \le m$ is the quality parameter of the original classical digital net, the so-called "t-value". We note that the t-value is in fact also dependent on m. From (7) it is clear that to effectively have convergence, one must have, at the very least,

$$m > t'_{\alpha s} + \lfloor s(\alpha-1)/2 \rfloor . \quad (8)$$

Following from this we now point out that the applicability of higher order nets is severely limited by the finite precision of the computing device. From (6) we need a precision of $n = \alpha m$ base b digits to represent the cubature nodes. Assuming base 2 and standard IEEE double precision numbers with 53 bits of precision, we are limited by how many points, $N = b^m$, we can represent for a given smoothness α. Using the best possible known $t'_{\alpha s}$ values, those from the Niederreiter-Xing points, see [17], we can calculate the minimum m needed for the convergence to kick in. We obtain the following tables:

α	maximum $N = 2^m$
2	2^{26}
3	2^{17}
4	2^{13}
5	2^{10}
6	2^{8}
7	2^{7}

$\alpha \setminus s$	1	2	3	4	5	6	7	$\cdots$	13	14
2	2	3	4	7	8	11	13	$\cdots$	25	(29)
3	3	6	9	13	16	(20)	(25)			
4	3	9	13	(18)	(25)	(30)	–			
5	5	(12)	(18)	(26)	–	–	–			
6	6	(14)	(22)	(31)	–	–	–			
7	(8)	(18)	(28)	–	–	–	–			

The numbers in brackets in the table on the right denote that such a level of smoothness in that many dimensions are impossible as we are running out of precision, cf. the table on the left, to represent the cubature nodes. We end this section with an overview of positive and negative points. Also here the focus of the negative points is on the dimensionality.

- To be able to represent the cubature nodes one needs a precision of $n = \alpha m$ base b digits. This might be prohibitive. It is easy to see that one quickly needs to resort to multi-precision calculations.
- The theoretical formula for the convergence (7) shows that quite a large number of cubature nodes is needed before the higher order convergence can be seen, cf. (8). In combination with the previous item this limits the applicability.
- It is not enough to construct the points in high precision. Also the integrand function itself needs to be evaluated in high precision. (For periodization this is less the case, as [13] shows.)

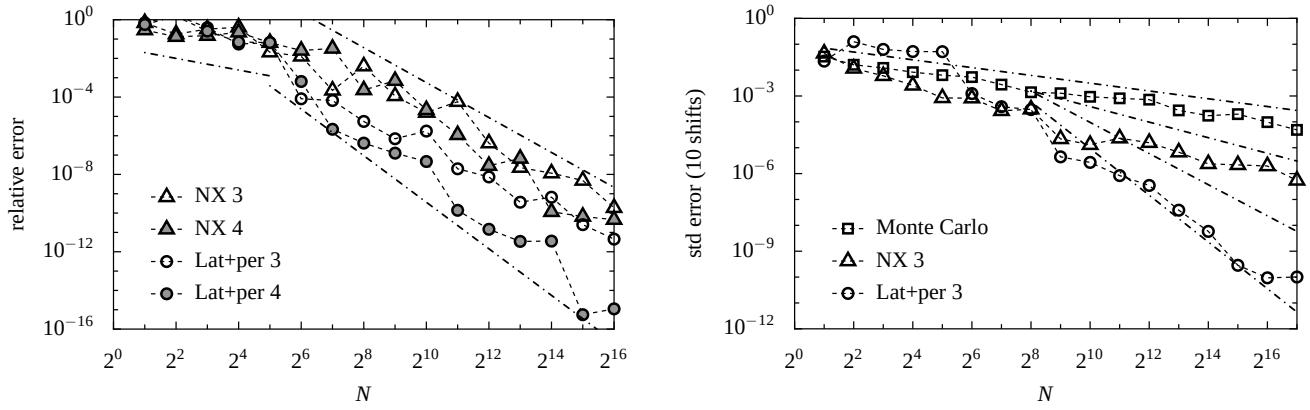

FIGURE 1. Lattice rule with periodization versus explicit higher order construction

+ Higher order nets still are equal-weight rules. This implies that the rule (1) is perfectly well conditioned, i.e., $\sum_{k=0}^{N-1} |w^{(k)}| = 1$, and thus this gives a numerically stable algorithm independent of the dimension. Of course we have just pointed out other reasons to increase precision when the dimension and/or smoothness increases.
+ The construction algorithm from [16] is explicit and it is easy to obtain the higher order generating matrices. In fact, in base 2, there is nearly no additional cost in generating points in comparison to classical nets (requiring no change; as long as the points fit in a supported word length). A lot of (classical) generating matrices are available, see [17].
+ The theoretical error bound has a much nicer form than that of periodization (if one has sufficient precision and time, i.e., large N).

A NUMERICAL COMPARISON

The following two examples were calculated in IEEE double precision (i.e., 53 bits of precision). For the lattice rule results we used a lattice sequence (which is an embedded series of lattice rules, see [7]) with the Sidi $\sin^m$-periodization [9] and generating vector taken from [14]. The higher order digital net results are based on a Niederreiter-Xing (NX) digital sequence in base 2, with generating matrices taken from [18].

First we present a toy example. On the left hand side of Figure 1 the error for integrating the 2-dimensional toy function $f(x_1,x_2) = x_1^4 x_2^3$ is plotted for the lattice sequence and the higher order sequence, both constructed to achieve $O(N^{-3})$ and $O(N^{-4})$. For reference we plotted dash-dotted lines of $O(N^{-1})$, $O(N^{-3})$ and $O(N^{-4})$. It can be seen that both QMC methods achieve the desired $O(N^{-3})$ result, but for the higher order digital net it is unclear if the $O(N^{-4})$ is achieved (its error for $O(N^{-3})$ and $O(N^{-4})$ stay close together). The higher order sequence does seem to show $O(N^{-4})$ from $N = 2^9$ up to 2^{12}, but overall this is hard to distinguish from the $O(N^{-3})$ result. One has to be careful due to the erratic convergence typical for a QMC method. However, the lattice sequence with periodization is clearly the better method here, achieving the promised order of convergence in both cases and outperforming the higher order net.

On the right hand side of Figure 1 we plot the error for integrating a 5-dimensional normal distribution up to some specified limits, cf. the application in [15], to achieve $O(N^{-3})$. The actual integrand is then 4-dimensional using the technique from [19]. For this function we also ran a Monte Carlo simulation which shows its expected $O(N^{-1/2})$ convergence. For reference we plotted dash-dotted lines of $O(N^{-1/2})$, $O(N^{-1})$, $O(N^{-2})$ and $O(N^{-3})$. In this case the higher order digital net construction does not reach the desired convergence. In comparison, the lattice sequence with periodization nicely gets $O(N^{-3})$ but has a visible bump for small N, doing worse than Monte Carlo in that range. This was already pointed out in our discussion as one of the negative side effects of periodization.

Similar results were obtained for other comparable test functions and more experiments with this normal integral with varying dimensions. For the latter the bump effect gets more and more pronounced for increasing dimensions, making the method unusable for the multivariate normal integral from around dimension 8 and further. In 8 dimensions one needs more than approximately 10^4 points to do better than plain quasi-Monte Carlo (or higher order digital nets), and in 9 dimensions the periodization results are always worse up to even 10^6 points.

CONCLUSION

In the eyes of the practitioner it might have seemed that the periodization strategy is doomed to failure, or only of minor practical value, moreover so, since the new techniques in [16] don't need periodization. A footnote in that style appeared in [12]. In the previous section however, we have shown that periodization still earns its place and in fact is the method of choice for the limited examples under consideration there.

However, higher order nets using the explicit construction come very close in our low-dimensional tests. It is probably interesting to keep a close eye on further developments of higher order digital nets as they clearly show some theoretical advantages over periodization. Furthermore, it is a trivial exercise to come up with a function for which higher order nets give very good results, while lattice rules with periodization would have trouble to cope (e.g., the constant function is integrated exactly by higher order nets). — So, should one use periodization? It depends: on the integrand, on the machinery and on the user.

ACKNOWLEDGMENTS

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State Science Policy Office. The scientific responsibility rests with its authors. The first author is a postdoctoral fellow of the Research Foundation Flanders (FWO). Part of this research was done while the first author was also a research associate at the University of New South Wales, Sydney, Australia.

REFERENCES

1. H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, Regional Conference Series in Applied Mathematics 63, SIAM, 1992.
2. I. H. Sloan, and S. Joe, *Lattice Methods for Multiple Integration*, Oxford Science Publications, 1994.
3. N. M. Korobov, *Number-Theoretic Methods in Approximate Analysis*, Goz. Izdat. Fiz.-Math., 1963, in Russian. English translation of results on optimal coefficients in [4].
4. A. H. Stroud, *Approximate Calculation of Multiple Integrals*, Automatic Computation, Prentice-Hall, 1971, ISBN 13-043893-6.
5. S. K. Zaremba, "La Méthode des "Bons Treillis" pour le Calcul des Intégrales Multiples," in *Applications of Number Theory to Numerical Analysis*, edited by S. K. Zaremba, Academic Press, 1972, pp. 39–119.
6. D. Nuyens, and R. Cools, Fast component-by-component construction of rank-1 lattice rules with a non-prime number of points, *J. Complexity* **22**, 4–28 (2006).
7. R. Cools, F. Y. Kuo, and D. Nuyens, Constructing embedded lattice rules for multivariate integration, *SIAM J. Sci. Comput.* **28**, 2162–2188 (2006).
8. F. Y. Kuo, and I. H. Sloan, Lifting the curse of dimensionality, *Notices Amer. Math. Soc.* **52**, 1320–1328 (2005).
9. A. Sidi, "A New Variable Transformation for Numerical Integration," in *Numerical integration IV (Oberwolfach, 1992)*, edited by H. Brass, and G. Hämmerlin, Birkhäuser Verlag, 1993, pp. 359–373.
10. D. P. Laurie, Periodizing transformations for numerical integration, *J. Comput. Appl. Math.* **66**, 337–344 (1996).
11. F. J. Hickernell, "Obtaining $O(n^{-2+\varepsilon})$ Convergence for Lattice Quadrature Rules," in *Monte Carlo and Quasi-Monte Carlo Methods 2000*, edited by K. T. Fang, F. J. Hickernell, and H. Niederreiter, Springer-Verlag, 2002, pp. 274–289.
12. F. Y. Kuo, I. H. Sloan, and H. Woźniakowski, Periodization strategy may fail in high dimensions, *Numer. Algorithms* **46**, 369–391 (2007).
13. I. Robinson, and M. Hill, Algorithm 816: r2d2lri: An algorithm for automatic two-dimensional cubature, *ACM Trans. Math. Software* **28**, 75–100 (2002), ISSN 0098-3500.
14. F. J. Hickernell, P. Kritzer, F. Y. Kuo, and D. Nuyens, Weighted compound integration rules with higher order convergence for all N (2010), working paper. `http://www.maths.unsw.edu.au/applied/pubs/appreprints2009.html`.
15. P. P. Boyle, T. Lai, and K. S. Tan, Pricing options using lattice rules, *North Amer. Act. J.* pp. 50–76 (2005).
16. J. Dick, Walsh spaces containing smooth functions and quasi-Monte Carlo rules of arbitrary high order, *SIAM J. Numer. Anal.* **46**, 1519–1553 (2008).
17. R. Schürer, and W. C. Schmid, "MINT: A Database for Optimal Net Parameters," in *Monte Carlo and Quasi-Monte Carlo Methods 2004*, edited by H. Niederreiter, and D. Talay, Springer-Verlag, 2006, pp. 457–469.
18. G. Pirsic, Generator matrices of NX-sequences, `http://www.ricam.oeaw.ac.at/people/page/pirsic/niedxing/` (2010).
19. A. Genz, Numerical computation of multivariate normal probabilities, *J. Comp. Graph Stat.* **1**, 141–149 (1992).